



Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Computerized segmentation and measurement of chronic wound images

Mohammad Faizal Ahmad Fauzi ^{a,*}, Ibrahim Khansa ^b, Karen Catignani ^c, Gayle Gordillo ^{b,e}, Chandan K. Sen ^{d,e}, Metin N. Gurcan ^f

^a Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, Malaysia

^b Department of Plastic Surgery, The Ohio State University, Wexner Medical Center, Columbus, OH, USA

^c Department of Internal Medicine, The Ohio State University, Wexner Medical Center, Columbus, OH, USA

^d Department of Surgery, The Ohio State University, Wexner Medical Center, Columbus, OH, USA

^e OSU Comprehensive Wound Center, Center for Regenerative Medicine and Cell-Based Therapies, The Ohio State University, Wexner Medical Center, Columbus, OH, USA

^f Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA

ARTICLE INFO

Article history:

Received 12 December 2014

Accepted 19 February 2015

Keywords:

Wound segmentation

Wound measurement

Wound analysis

Chronic wound

Probability map

ABSTRACT

An estimated 6.5 million patients in the United States are affected by chronic wounds, with more than US \$25 billion and countless hours spent annually for all aspects of chronic wound care. There is a need for an intelligent software tool to analyze wound images, characterize wound tissue composition, measure wound size, and monitor changes in wound between visits. Performed manually, this process is very time-consuming and subject to intra- and inter-reader variability. In this work, our objective is to develop methods to segment, measure and characterize clinically presented chronic wounds from photographic images. The first step of our method is to generate a Red-Yellow-Black-White (RYKW) probability map, which then guides the segmentation process using either optimal thresholding or region growing. The red, yellow and black probability maps are designed to handle the granulation, slough and eschar tissues, respectively; while the white probability map is to detect the white label card for measurement calibration purposes. The innovative aspects of this work include defining a four-dimensional probability map specific to wound characteristics, a computationally efficient method to segment wound images utilizing the probability map, and auto-calibration of wound measurements using the content of the image. These methods were applied to 80 wound images, captured in a clinical setting at the Ohio State University Comprehensive Wound Center, with the ground truth independently generated by the consensus of at least two clinicians. While the mean inter-reader agreement between the readers varied between 67.4% and 84.3%, the computer achieved an average accuracy of 75.1%.

© 2015 Published by Elsevier Ltd.

1. Introduction

Computer-aided measurement of the size and characteristics of chronic wounds is a novel approach to standardizing the accuracy of chronic wound assessment. A chronic wound, as defined by Centers for Medicare and Medicaid Services, is a wound that has not healed in 30 days. An estimated 6.5 million patients in the United States are affected by chronic wounds, and it is claimed that an excess of US\$25 billion is spent annually on treatment of chronic wounds. The burden is growing rapidly due to increasing health care costs, an aging population and a sharp rise in the incidence of diabetes and obesity worldwide [1]. As such, there is a need for a

timely and accurate method to document the size and evolving nature of chronic wounds in both the inpatient and outpatient settings. Such an application can potentially reduce clinicians' workload considerably; make the treatment and care more consistent and accurate; increase the quality of documentation in the medical record and enable clinicians to achieve quality benchmarks for wound care as determined by the Center for Medicare Services.

The current state of the art approach in measuring wound size using digital images, known as digital planimetry, requires the clinician to identify wound borders and wound tissue type within the image. This is a time-intensive process and is a barrier to achieving clinical quality benchmarks. Our group is developing image analysis tools that will enable the computer to perform this analysis rather than requiring user input. Developing an accurate method of measuring wound size and tissue characteristics serially over time will yield clinically meaningful information in relation to

* Correspondence to: Faculty of Engineering, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Selangor, Malaysia. Tel.: +1 603 8312 5330.

E-mail address: faizal1@mmu.edu.my (M.F. Ahmad Fauzi).

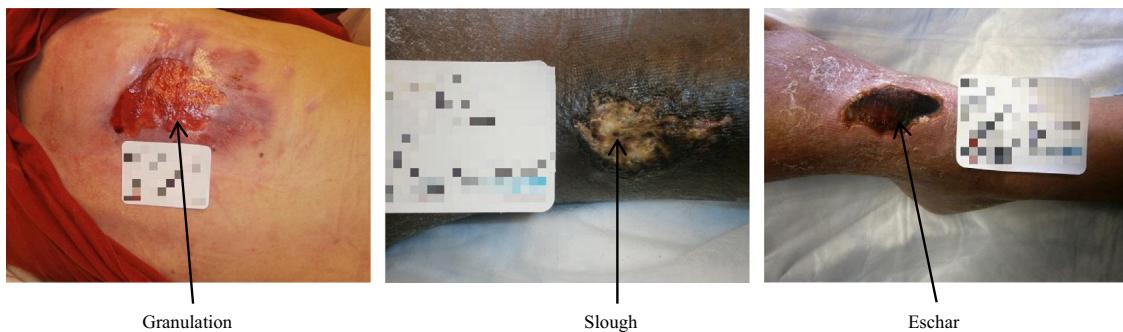


Fig. 1. Most commonly seen tissues in wounds: granulation, slough and eschar.

the progression or improvement of the wound. The focus of the work reported in this paper is the segmentation of the wounds.

A wound exhibits a complex structure and may contain many types of tissue such as granulation, slough, eschar, epithelialization, bone, tendon and blood vessels, each with different color and texture characteristics. In this paper, we proposed a novel probability map that measures the likelihood of wound pixels belonging to granulation, slough or eschar (see Fig. 1), which can then be segmented using any standard segmentation techniques. In this work, we focus on the granulation, slough and eschar tissues as these are the three most commonly seen tissues in wounds. A preliminary version of this work has been reported in [2]. This paper extended the previous work significantly with an extensive literature review, more elaborate explanation of the proposed method, employing two segmentation techniques to show that the probability map is adaptable to many different techniques, comparison with other existing method, comprehensive analyses on inter-reader variability between clinicians, a much bigger dataset used for performance evaluation (which was divided into three sets for analysis purpose), as well as more elaborate discussions on the results.

The paper is organized as follows: Section 2 presents the review of the literature on wound image analysis. In Section 3, we present our proposed probability map approach to wound segmentation and integrate it with two different segmentation techniques. Section 4 and 5 discusses the experimental setup, results and discussion. Finally, Section 6 concludes the paper and describes future work.

2. Literature review

Although wound segmentation from photographic images has been the subject of several studies, most of the work in this area deals with images that are either acquired under controlled imaging conditions [3–4], confined to wound region only [4–7], or narrowed to specific types of wounds [7–9]. Because these restrictions are mostly impractical for clinical conditions, there is a need to develop image segmentation methods that will work with images acquired in regular clinical conditions.

Table 1 summarizes current works in wound segmentation and monitoring as well as existing software tools. Wannous et al. [3] compared the mean shift, JSEG and CSC techniques in segmenting 25 wound images, before extracting color and textural features to classify the tissues into granulation, slough and necrosis using an SVM classifier. The wound images were taken with respect to a specific protocol integrating several points of views for each single wound, which includes using a ring flash with specific control and placing a calibrated Macbeth color checker pattern near the wounds. They reported that both segmentation and classification work better on granulation than slough and necrosis. Hettiarachchi et al. [4] attempted wound segmentation and measurement in a mobile

setting. The segmentation is based on active contour models which identifies the wound border irrespective of coloration and shape. The active contour process was modified by changing the energy calculation to minimize points sticking together as well as including pre-processing techniques to reduce errors from artifacts and lighting conditions. Although the accuracy was reported to be 90%, the method is rather sensitive to camera distance, angle and lighting conditions.

In the work by Veredas et al. [5], a hybrid approach based on neural networks and Bayesian classifiers is proposed in the design of a computational system for tissue identification and labeling in wound images. Mean shift and region-growing strategy are implemented for region segmentation. The neural network and Bayesian classifiers are then used to categorize the tissue based on color and texture features extracted from the segmented regions, with 78.7% sensitivity, 94.7% specificity and 91.5% accuracy reported. Hani et al. [6] presented an approach based on utilizing hemoglobin content in chronic ulcers as an image marker to detect the growth of granulation tissue. Independent Component Analysis is employed to extract grey level hemoglobin images from Red-Green-Blue (RGB) color images of chronic ulcers. Data clustering techniques are then implemented to classify and segment detected regions of granulation tissue from the extracted hemoglobin images. 88.2% sensitivity and 98.8% specificity were reported on a database of 30 images.

Perez et al. [7] proposed a method for the segmentation and analysis of leg ulcer tissues in color images. The segmentation is obtained through analysis of the red, green, blue, saturation and intensity channels of the image. The algorithm, however, requires the user to provide samples of the wound and the background before the segmentation can be carried out. Wantanajittikul et al. [8] employs the Cr-transformation, Luv-transformation and fuzzy c-means clustering technique to separate the burn wound area from healthy skin before applying mathematical morphology to reduce segmentation errors. To identify the degree of the burns, h-transformation and texture analysis are used to extract feature vectors for SVM classification. Positive predictive value and sensitivity between 72.0% and 98.0% were reported in segmenting burn areas in five images, with 75.0% classification accuracy.

Song and Sacan [9] proposed a system capable of automatic image segmentation and wound region identification. Several commonly used segmentation methods (*k*-means clustering, edge detection, thresholding, and region growing) are utilized to obtain a collection of candidate wound regions. Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) are then applied with supervised learning in the prediction procedure for the wound identification. Experiments on 92 images from 14 patients (78 training, 14 testing) showed that both MLP and RBF have decent efficiency, with their own advantages and disadvantages. Kolesnik and Fexa [10–12] used color and textural features from 3-D color histogram, local binary pattern and local contrast variation with the support vector machine (SVM) classifier to segment 23 wound images based on 50 manually segmented training images. The SVM generated wound boundary is further refined using

Table 1

Summary of current works on wound segmentation, monitoring and software tools.

Papers	Addressing	Segmentation/classification methodology	Image types	Number of images	Segmentation/classification accuracy
Wannous et al. [3]	Wound segmentation	Mean shift, JSEG, CSC & SVM	With background, controlled conditions	25	73.3–80.2% (granulation), 56.4–69.8% (slough), 64.9–70.7% (necrosis)
Hettiarachchi et al. [4]	Wound Segmentation	Active Contour	Wound region only, controlled conditions	20	90.0%
Veredas et al. [5]	Wound segmentation and tissue characterization	Mean shift & region growing, neural networks and Bayesian classifiers	Wound region only	113 divided into 10 testing sets	78.7% sensitivity 94.7% specificity 91.5% accuracy
Hani et al. [6]	Granulation detection and segmentation	ICA and k-means	Wound region only	30	88.2% sensitivity 98.8% specificity
Perez et al. [7]	Wound segmentation and analysis	RGBSI analysis, user need to provide samples of wound and background for each image	Wound region only, leg ulcers only	Not mentioned	Visual observations only
Wantanajittikul et al. [8]	Burn image segmentation and characterization	FCM & morphology, texture analysis and SVM	With background, but burn cases only	5	72.0–98.0% (segmentation) 75.0% (classification)
Song and Sacan [9]	Wound segmentation	Neural Networks, K-means Clustering, Edge Detection, Thresholding, and Region Growing	Foot ulcers only	78 training, 14 testing	71.4% (MLP) 85.7% (RBF)
Kolesnik and Fexa [10–12]	Wound segmentation	SVM, texture and deformable snake	With some background	50 training, 23 Testing	Error rate of 6.6% (color), 22.2% (texture), 5.8% (hybrid)
Cukjati et al. [13]	Wound healing rate measurement	Not applicable	Not applicable	Not applicable	Not applicable
Bums et al. [14]	Software to study healing rate for a given patient population	Not applicable	Foot ulcer only	Not applicable	Not applicable
Loizou et al. [15]	Wound healing monitoring	Snake (segmentation), texture feature (healing)	Wound region only, foot wounds only	40 images from 10 cases	Not available
PictZar [16]	Software for wound analysis	Manual drawing and calibration	With background	Not available	Not available
Filko et al. [17]	Software for wound analysis and healing monitoring	Not applicable	Cropped or hand-drawn region of wound images	6 images from 1 case	Not applicable
Weber et al. [18]	Hardware and software to capture wound mapping	Using electrode to obtain wound mapping and characteristics	Not applicable	Not applicable	Not applicable

deformable snake adjustment. Although this study does not have the aforementioned restrictions (i.e. acquired under controlled imaging conditions, confined to wound region only, or narrowed to specific types of wounds), results were reported on a relatively small set of images. An average error rate of 6.6%, 22.2% and 5.8% were reported for the color, texture and hybrid features, respectively.

In addition to wound segmentation, wound healing and monitoring have been the subject of several studies on wound image analysis. Cukjati et al. [13] presented their findings on how the wound healing rate should be defined to enable appropriate description of wound healing dynamics. They suggested that wound area measurements should be transformed to percentage of initial wound area and fitted to a delayed exponential model. In the suggested model, the wound healing rate is described by the slope of the curve is fitted to the normalized wound area measurements over time after initialization delay. Loizou et al. [14] established a standardized and objective technique to assess the progress of wound healing in a foot. They concluded that while none of the geometrical features (area, perimeter, x-, y-coordinate) show significant changes between visits, several texture features (mean, contrast, entropy, SSV, sum variance, sum average) do, indicating these features might provide a better wound healing rate indication. Finally, Burns et al. [15] evaluated several methods for quantitative wound assessment on diabetic foot ulcers, namely wound volume, wound area, and wound coloration.

There are also quite a few software tools for wound analysis and monitoring currently available. All the software; however, has yet to incorporate automated or semi-automated wound detection or

segmentation so that the clinician's initial involvement can be minimized. For example, PictZar Digital Planimetry Software [16] is commercial software for wound analysis which provides measurements such as length, width, surface area, circumference, and estimated volume to the users. The software, however, does not incorporate automated or semi-automated wound detection; instead it requires user drawings and calibration for the above measurements to be computed. Filko et al. [17] developed WITA, a color image processing software application that has the capability to analyze digital wound images, and based on learned tissue samples, the program classifies the tissue and monitors wound healing. The wound tissue types are divided into black necrotic eschar, yellow fibrin or slough, red granulation tissue and unclassified parts of the image, although no evaluation against the known ground truth was presented for the image analysis part of the software. To obtain wound dimensions, users must mark the distance on the photograph that is equivalent to 1 cm (or 1 in.). A different approach to wound monitoring software and hardware was proposed by Weber et al. [18]. They developed a new "wound mapping" device, which is based on electrical impedance spectroscopy and involves the multi-frequency characterization of the electrical properties of wound tissue under an electrode array. This approach, however, requires major changes to the daily clinical routine in wound care.

3. Wound segmentation based on a probability map

The wound images used in our experiments are provided by the Comprehensive Wound Center of the Ohio State University Wexner



Fig. 2. Example of wound images with complicated backgrounds used in the experiment.

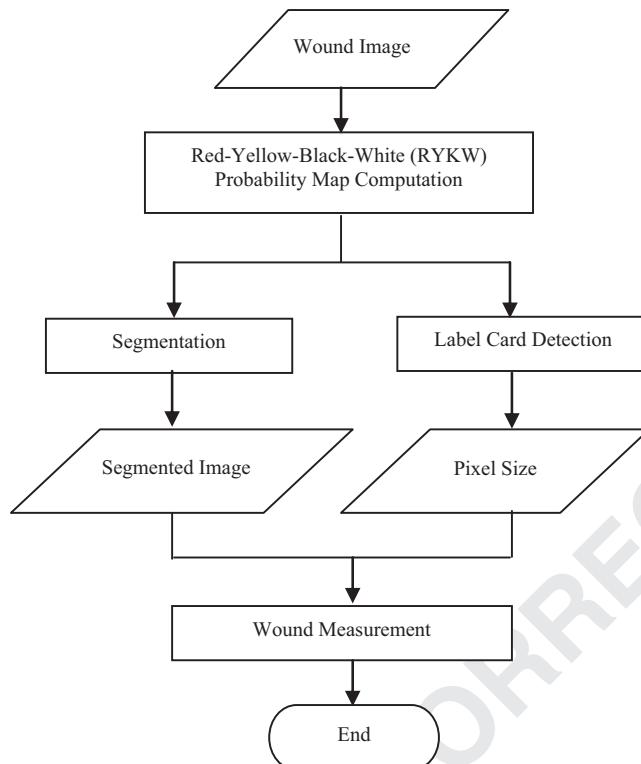


Fig. 3. Flowchart of the proposed method.

Medical Center, with Institutional Review Board (IRB) approval. The center is one of the largest wound centers in the US, and the wound images captured in the center comes from different camera manufacturers, setting and capture conditions: different medical center employees (not professional photographers) capturing the images in routine clinical work using different cameras. This simulates the variation that we expect to see in other medical centers in terms of patient variability as well as variation due to image capture. Unlike the wound images used in the literature [3–18], these images present additional challenges. As discussed in the previous section, many previous works in this field are typically carried out in regions that contain the wound only, thus they do not have to deal with the issue of complicated background, especially those red, yellow and black objects, interfering with the segmentation process (see Fig. 2). In order to simplify the task at this stage, the algorithm requires the user to mark a single point (i.e. a single click) inside the wound to start the segmentation process.

Our proposed method consists of several stages as shown in Fig. 3. The first step is the red-yellow-black-white (RYKW) probability map computation in a modified HSV (Hue-Saturation-Value) color space (Section 3A). Once the probability map is established, the next step is the segmentation of the boundaries of the wound in the area (Section 3B). We present the results of two different

segmentation approaches: region growing segmentation and optimal thresholding. Because the distance between the camera and the wound is not recorded, this information needs to be extracted by the content in the image. We developed a novel approach, which analyzes the image to detect patient labels, typically attached near the wound and uses the size of the label to calibrate the wound size measurements (Section 3C).

3.1. Probability map computation

Granulation, slough and eschar tissues generally correspond to red (R), yellow (Y) and black (K) tissues, respectively in the wound area (see Fig. 1). Due to the fact that the subsequent stage requires the detection of white label cards, as well as to avoid any white pixels in the image being wrongly classified as yellow, a fourth color, white (W) is included in the probability map computation, resulting in a four-dimensional (4D) RYKW map. Given a wound image, our method computes the probability of each pixel in the image belonging to one of these colors. The probability is computed based on the distance of the image pixels to the red, yellow, black and white colors in a modified HSV color space. The HSV color space was chosen because it can be modified to maximize the distances between the four colors of interest (refer to Eq. 2 and Eq. 3).

Consider an image I , probability matrix P , and color set $C_k = \{R, Y, K, W\}$ where $k = 1, 2, 3, 4$ represents the 4 colors R, Y, K, W respectively. For a particular pixel x within I , the probability p of the pixel belonging to a color C_k (i.e. one of red, yellow, black or white) is computed through the following equation:

$$p_k(x) = \frac{1}{\left(\frac{d(C_k, x)}{d(R, x)}\right)^2 + \left(\frac{d(C_k, x)}{d(Y, x)}\right)^2 + \left(\frac{d(C_k, x)}{d(K, x)}\right)^2 + \left(\frac{d(C_k, x)}{d(W, x)}\right)^2} \quad (1)$$

where $d(C_k, x)$ is the distance (see Eqs. (4–7)) between the value of pixel x and the particular color C_k . In other words, the probability is inversely proportional to the relative distance between the pixel and the color of interest. The above equation is applied to all pixels for all four colors, producing a 4D probability map, P , with the sum of the probability at any one pixel is equal to 1. The probability definition used here is similar to that of the fuzzy c-means clustering method without the fuzzifier parameter [19]. From the image point of view, the 4D matrix P can be viewed as a stack of 4 single matrices P_k , each showing the probability of the wound image pixels belonging to the 4 different colors. From the pixel point of view, the matrix P can be viewed as a collection of many vectors p , each showing the probability of individual pixels belonging to the 4 colors of interest.

One of the challenges in wound segmentation is to differentiate between regions with similar hue characteristics: e.g. dark red (granulation) vs. black (eschar) regions, as well as light yellow (slough) vs. white (epibole, skin etc.) regions. Fig. 4 shows an example of a dark red granulation tissue whose Value channel, V , values range between 0.2 and 0.4. Taking $V = 0.5$ as the threshold, the tissue would have been misclassified as being closer to black rather than red (where 0 refers to pure black, and 1 refers to pure red). This, combined with the close proximity between red and

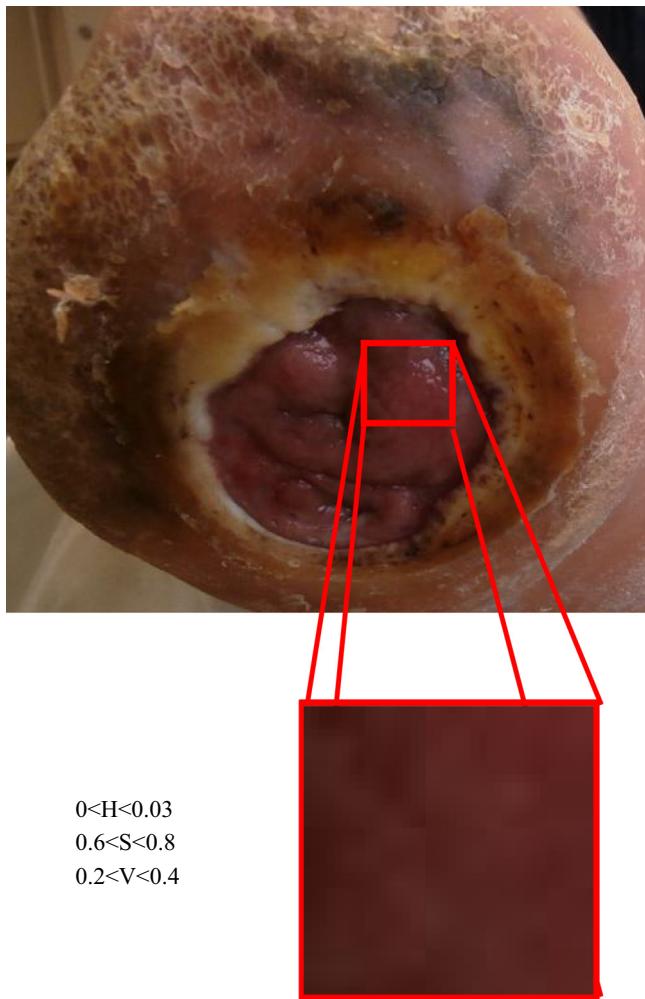


Fig. 4. Illustration of red granulation tissue mistaken as black eschar tissue.

yellow colors, makes segmentation of the three tissue types complicated, regardless of the color model used (RGB, HSV, CIE L*a*b* etc.). In this work, we developed a modified HSV color model to improve the accuracy of the probability map by scaling the Saturation (S) and Value (V) components according to Eqs. 2 and 3, respectively to obtain S_{mod} and V_{mod} :

$$S_{mod} = \frac{\log(\alpha*S+1)}{\log(\alpha+1)} \quad (2)$$

$$V_{mod} = \frac{\log(\alpha*V+1)}{\log(\alpha+1)} \quad (3)$$

where S_{mod} and V_{mod} are the modified Saturation and modified Value respectively, and α is a constant. In our work, we have chosen $\alpha = 8$ so that the first quarter of the original scale (dark or light regions) will be stretched to half the modified scale, while the remaining three quarters of the original scale (red or yellow regions) will be compressed to occupy the remaining half of the modified scale (see Fig. 5). Furthermore, the Hue (H) component is also shifted by 30° to maximize the distance between red and yellow.

Fig. 5(a) shows the transformation of S and V using Eq. (2) and Eq. (3), while Fig. 5(b) and (c) shows the transformation of the black-red, black-yellow, white-yellow and red-yellow color transition from the standard HSV to the modified HSV color model. It can be observed that the modified HSV model better reflects the color distances between the four colors of interest. Under the standard HSV, dark red and dark yellow are closer to black; similarly light red and light yellow are closer to white. This

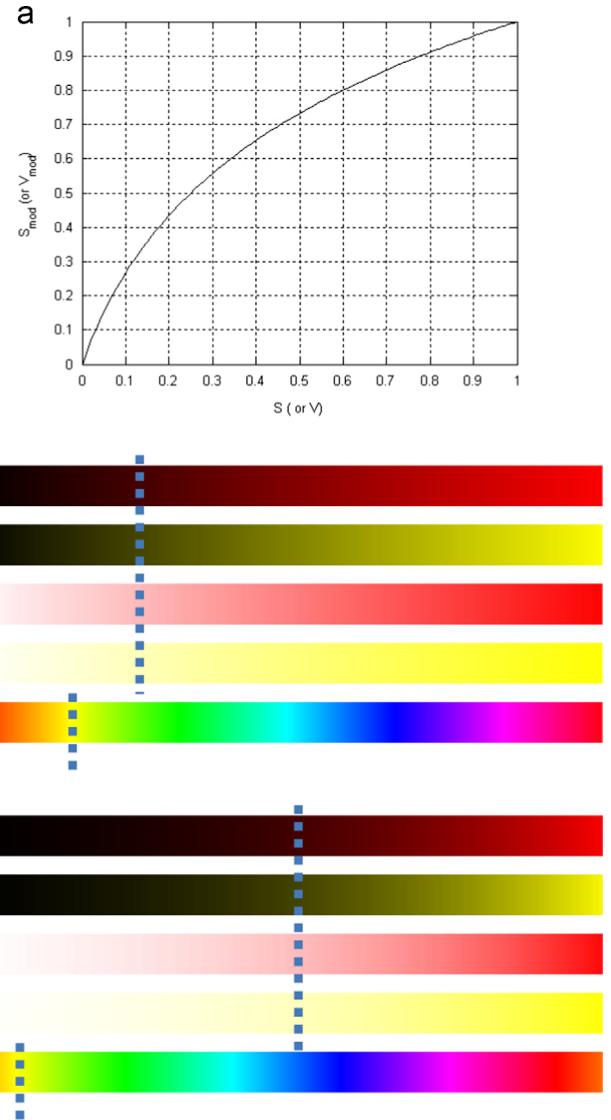


Fig. 5. (a) Transformation of S and V to S_{mod} and V_{mod} , based on Eqs. (2) and (3) ($\alpha=8$), and color transitions for (b) standard HSV, and (c) modified HSV color models. The dashed lines show how the H, S and V values are shifted after the transformation.

would negatively affect the accuracy of the color probability map. The proposed modified HSV model is thus better suited to computing the probability of pixels belonging to any one of the four colors (see Section 5B and Table 14 for comparison between the modified and original HSV for region growing).

Due to the uneven color distribution of the HSV or modified HSV color models (e.g. dark colors occupied almost half of the entire color space), the calculation of distance, $d(C_k, x)$ between a particular pixel, x and the colors is defined differently for the four colors. The distance of a pixel to black is based solely on V_{mod} , while the distance to white is based on V_{mod} and S_{mod} . The distances to red and yellow on the other hand make use of all V_{mod} , S_{mod} and H_{mod} . For a particular pixel x , the proposed distance equations are summarized below:

$$d(R, x) = \sqrt{(H_{mod}(x) - H_{mod}(R))^2 + (S_{mod}(x) - S_{mod}(R))^2 + (V_{mod}(x) - V_{mod}(R))^2} \quad (4)$$

$$d(Y, x) = \sqrt{(H_{mod}(x) - H_{mod}(Y))^2 + (S_{mod}(x) - S_{mod}(Y))^2 + (V_{mod}(x) - V_{mod}(Y))^2} \quad (5)$$

$$d(K, x) = V_{mod}(x) - V_{mod}(K) \quad (6)$$

$$d(W, x) = \sqrt{(V_{mod}(x) - V_{mod}(W))^2 + (S_{mod}(x) - S_{mod}(W))^2} \quad (7)$$

where the following values are defined:

$$V_{mod}(K) = 0 \quad V_{mod}(W) = 1 \quad S_{mod}(W) = 0$$

$$H_{mod}(R) = 11/12 \quad S_{mod}(R) = 1 \quad V_{mod}(R) = 1$$

$$H_{mod}(Y) = 1/12 \quad S_{mod}(Y) = 1 \quad V_{mod}(Y) = 1$$

3.2. Segmentation

While there are many possible segmentation methods for use in medical applications, e.g. [20–25], we based our segmentation on two well-known and rather basic techniques, namely region growing segmentation, and optimal thresholding. We will demonstrate that even with these two simple segmentation algorithms, when coupled with our proposed probability map approach, is able to provide reliable segmentation of wounds. While the proposed approach works with the selection of an initial seed point by a clinician, the RYKW map has the potential to improve the segmentation into fully automated segmentation. This can be achieved by first identifying all potential wound regions throughout the entire image based on color information, before carrying out advanced image analysis to filter the false positive regions, leaving only true wound regions as the segmented output.

3.2.1. Region growing

Region growing [26] is a pixel-based image segmentation algorithm that examines neighboring pixels of initial seed points and determines whether neighbors of the pixel should be added to the region. In our proposed method, the initial seed points are to be provided by the clinician. The regions are grown from the initial seed point's probability vector to adjacent points based on the 4D probability map P (Eq. 1). A neighbor is added to the region if the distance between that pixel's probability vector and the mean probability vector of the region (i.e. the mean probability of each R, Y, K, W channel over the current segmented region) is less than a certain threshold value, t . The process continues until either all the neighbor's distances are above the threshold, or all the pixels have been processed.

To ensure the region growing process does not stop prematurely, a mechanism is included to search for pixels with a similar probability map within a certain radius, r , from the region boundary, and the process continues. Morphological closing and filling operations are then applied during post-processing to remove noise and soften the edges. From experiments, suitable values for the threshold and radius are $t=0.1$ and $r=5$ pixels, respectively. Note that the proposed algorithm only segments the granulation, slough and/or eschar regions and ignores the rest of the image as clinicians are only interested in the wounds. While region growing is generally considered as computationally expensive operation, the probability map really helps to speed up the process by providing a valuable color discriminator between the four colors of interest.

3.2.2. Optimal thresholding

Our “optimal thresholding approach” segments the image by thresholding the difference matrix of the probability map, P , while taking into account the pixel's tissue type and strength of its probability. While there are many available thresholding methods such as Otsu thresholding that can be used to segment the probability map, these methods are rather “hard” thresholding methods; if single wounds are inadvertently separated to 2 or more smaller wounds (which can happen very frequently due to

illumination etc.), the segmentation can be considered to fail since the calculated accuracy (refer Section 4) will be very low.

The key idea behind our approach is first to identify all pixels whose color characteristics are similar to those of the seed pixel, before iteratively refining the segmentation boundary. The refinement is by simple thresholding of the difference matrix, Q , which is a matrix of the difference between the two highest probabilities for each pixel, and provides a second degree of tissue membership probability:

$$Q = P_{max1} - P_{max2} \quad (8)$$

where $P_{max1} = \max(P)$ and $P_{max2} = \max(P)|_{P \neq P_{max1}}$

Given the seed point pixel and its probability vector, its highest probability tissue class is identified, and pixels with the following properties are considered for segmentation:

Property 1. Pixels with the same tissue class as their highest probability. Value of Q ranges from 0 to the maximum value in Q , φ .

Property 2. Pixels with the same tissue class as their second highest probability, and in which their difference with the highest probability is below a certain threshold, τ . Value of Q ranges from 0 to $-\tau$.

In the strictest sense, only pixels with **Property 1** should be included in the segmented region; however, due to the complicated nature of the wound tissue, pixels with **Property 2** are also included to minimize false negative. The region of interest (ROI) at this point is defined as the region in Q whose pixels satisfy either **Property 1** or **Property 2**, with values ranging between the φ and $-\tau$.

The next step is to iteratively threshold the ROI, starting from φ . At each step the mean of the segmented ROI where the seed point is located is calculated. Theoretically the mean will decrease as the threshold value decreases towards $-\tau$. The *optimal threshold* is defined as the threshold value where the mean values become ‘stable’ without any sudden decreases or increase. The segmented wound region can then be obtained by thresholding the ROI with the optimal threshold value. As in the region growing, morphological closing and filling operations are then applied during post-processing to obtain the final segmentation. Experimentally, the suitable values for the threshold, τ , and step size decrement, $step$, are $\tau = 0.1$ and $step = 0.01$, respectively. The whole process is summarized as pseudo-code in Table 2.

3.2.3. Label card detection and wound measurement

Since the distance between the camera and the wound is not recorded, the absolute values for wound measurements – necessary for clinical reporting – cannot be recorded. To solve this problem, we have developed a technique to automatically scale the

Table 2

Pseudo-code of the optimal-thresholding based segmentation approach.

Input: 4D probability map, P

Output: Segmented wound region, I_{seg}

Procedure:

- 1 Compute probability difference matrix, Q
- 2 Based on probability map of seed pixel, identify *ROI*
- 3 Set $\varphi = \max(Q)$
- 4 Set $\tau = 0.1$
- 5 Set $step = 0.01$
- 6 Set $th = \varphi$
- 7 While $th > -\tau$
 - $seg = ROI > th$
 - $segmean = mean(seg)$
 - $th = th - step$
- 8 Identify optimal threshold, th_{opt} based on $segmean$
- 9 $I_{seg} = ROI > th_{opt}$
- 10 Perform morphological operations on I_{seg}

wound size. As in most medical centers, each of the wound images taken at the Wexner Medical Center contains a white label card, which we automatically detected and used as a reference to compute the actual pixel size in the image. The white label card has a standard size of 4.5 cm by 6.5 cm. With successful detection of the card and its size with respect to the image, we can easily calculate the pixel measurements in cm per pixel unit.

To detect the card, first the white regions are identified from the same RYKW map computed in the previous step ([Section 3A](#)). Then, the detected white regions are filtered based on their area, rectangularity (actual area over minimum bounding rectangle area) and convexity measure (actual area over convex hull area) to identify potential rectangular regions for the white card. The rectangularity and convexity measure helps in eliminating irregular shape region, while the area relative to the image size helps in eliminating false rectangular regions. The length and width of the identified label card are then used to calibrate the pixel size. With the pixel size available, measuring the wound size is straightforward. Currently, the proposed algorithm outputs three measurements: area, length (major diameter) and width (minor diameter) of the segmented wound.

4. Experimental setup

This study was done with the institutional review board (IRB) approval. In our experiments, we used a total of 80 images, whose ground truth was provided by at least two clinicians. The images are of 768×1024 pixels in resolution, stored in JPEG format. They were captured by the clinicians following normal clinical practice and under non-controlled conditions, i.e. no measures were taken to control the illumination, background or the wound to background ratio, resulting in a very challenging set of images. To capture the ground truth, an in-house software tool was developed. Using this tool, clinicians can not only draw the boundaries of the wound but also its three tissue components: granulation, slough and eschar tissues. Again using this tool, the user can input the estimates (as a percentage) for tissue components that are already an integral part of wound documentation. The clinicians first manually drew the wound boundaries for each image independently. Based on the drawn boundaries, the clinicians were then asked to estimate the percentage of granulation, slough and eschar tissues before proceeding to draw the boundaries for each tissue type. The tool is capable of handling as many number of wound or tissue regions possible, hence the clinicians were asked to provide as detail a drawing as possible. Depending on the complexity of the image, clinicians spent between 30 s and 3 min to annotate a single image.

The images were divided into three sets as shown in [Table 3](#). Set 1, consisting of 10 images, were annotated with the consensus of three clinicians, and used as a training set to ensure that all three clinicians have the same understanding in defining the different tissue types as well as their boundaries. Set 2, with 15 images, were annotated by all three clinicians separately, producing three separate ground truth files for each image. Finally Set 3, with 55 images, were annotated by two clinicians independently, resulting in two separate ground truth files. The wound and tissue boundaries from the ground truth files of Sets 2 and 3 are compared to evaluate the level of agreement between the clinicians. Tissue component percentage estimation by the clinicians were also compared to the actual tissue percentage from the drawings to evaluate the accuracy of the clinicians' estimation.

The inter-reader variability is measured using the agreement measure in Eq. (9):

$$\text{Agreement} = \frac{D_1 \cap D_2}{D_1 \cup D_2} D_2 \times 100 \quad (9)$$

where D_1 and D_2 refer to the region annotated by the first, second

Table 3

Categorization of images.

Sets	Number of images	Number of ground truth
Set 1	10	1 (consensus from 3 clinicians)
Set 2	15	3 (from 3 clinicians)
Set 3	55	2 (from 2 clinicians)

Table 4

wound agreement for set 2 images (percentage accuracy measure in Eq. (9)).

Agreement Between	Mean	Min	Max	Med	Std Dev
Clinicians 1,2 and 3	74.3	40.7	88.3	76.3	12.5
Clinicians 1 and 2	84.3	69.7	94.4	86.1	7.4
Clinicians 1 and 3	81.5	41.4	92.3	87.4	13.0
Clinicians 2 and 3	80.3	55.0	92.7	83.2	10.5

or third clinician, respectively. Due to the high degree of inter-reader variability (to be discussed in [Section 5](#)), it is difficult to obtain one common ground truth for Sets 2 and 3. Hence, to evaluate the accuracy of computer segmentation, the resulting segmentation is compared to each of the different ground truths. In other words, the segmentation results are compared to each clinician's manual drawings, thereby indicating with which proposed algorithm the clinicians tend to agree more.

The same measurement in Eq. (9) is used to determine the accuracy of the computer segmented regions against the ground truth:

$$\text{Accuracy} = \frac{GT \cap CS}{GT \cup CS} \times 100 \quad (10)$$

where GT refers to the boundaries drawn by any one of the clinicians, and CS refers to the computer segmented region.

5. Experimental results and discussion

We first present the inter-reader variability between clinicians on the wound boundaries, tissue characterization as well as tissue percentage estimation in [Section 5A](#). The proceeding sub-section will then report the results of the computer segmentation against all the ground truth discussed in [Section 5B](#).

5.1. Inter-reader variability between clinicians

As explained in [Section 4](#), three clinicians independently drew the boundaries of the wounds in Set 2 as well as estimated the percentages of tissue types. In this section, this data will be used to evaluate inter-reader variability. [Table 4](#) shows the statistics of wound boundary agreement between the clinicians for the images in Set 2. Since there are three clinicians involved, four sets of comparison are carried out. As can be observed from [Table 4](#), the mean agreement between any two clinicians varies between 80.3% and 84.3%. The mean drops to 74.3% when all three clinicians' readings are compared, indicating that it is more difficult to reach an agreement when more clinicians involved (the trend for the median agreement follows a similar trend). Note that the minimum agreement goes as low as 40.7%, which suggests that some of the wounds are quite complicated and thus their boundaries are relatively difficult to define.

[Table 5](#) shows the statistics for images in Set 3. Clearly, with more images, the mean and median agreement between clinicians 2 and 3 (clinician 1 is not involved in evaluating Set 3) drops rather sharply, from 80.3% to around 67.4% in mean agreement, and from 83.3% to 70.8% in median. The standard deviation also almost doubles, while the minimum agreement can be as low as 24.4%. This suggests that with increased number of images to annotate,

some of which contain relatively complicated wounds, the agreement between the clinicians plummets. This is another reason why we will be comparing the computer segmentation with the ground truth from individual clinicians instead of a combined ground.

To gauge intra-reader variability, we have also asked two of the clinicians to re-draw the wound boundary for a subset of cases (10 images) after a month from their initial reading. The intra-reader variability is summarized in [Table 6](#). As in the inter-reader variability ([Table 4](#)), the difference between two consecutive readings is relatively high, with average self-agreement of 80.4% and 84.5% for the two clinicians.

While the average agreement between the clinicians at the wound level may still acceptable, their agreement at the tissue level is much lower. [Tables 7 and 8](#) show the tissue characterization agreement between the clinicians for Sets 2 and 3, respectively. It can be seen that the mean and median agreement are all below 60% with standard deviation of mostly more than 30%. There were many instances where the clinicians do not agree on the particular tissue types within the wound, especially when it comes to differentiating granulation and slough, or between slough and eschar, and even granulation and epithelium. This is the reason for minimum agreement (all the values in the 'Min' column in [Tables 7 and 8](#)) to be 0%. In other words, there are always situations where one clinician will identify a particular region within the wound, with which the other clinician will not agree. For example, [Fig. 6](#) shows two examples of images with the lowest agreement between two clinicians. While the

Table 5
Wound agreement for Set 3 images (Percentage accuracy measure in Eq. (9)).

Agreement between	Mean	Min	Max	Med	Std dev
Clinicians 2 and 3	67.4	24.5	94.5	70.8	19.5

Table 6
Intra-reader variability for wound agreement.

Agreement between	Mean	Min	Max	Med	Std dev
Clinicians 2	84.5	66.0	97.3	87.2	10.5
Clinicians 3	80.4	58.0	97.3	84.8	14.7

Table 7
Tissue agreement for Set 2 images.

Tissue types	Agreement between clinicians	Mean	Min	Max	Med	Std dev	of img
Granul	1,2 and 3	42.9	0.0	86.2	42.4	31.6	19
	1 and 2	59.6	0.0	94.0	71.9	30.1	19
	1 and 3	50.9	0.0	89.2	54.2	35.0	19
	2 and 3	52.6	0.0	88.5	60.3	31.8	18
Slough	1,2 and 3	17.8	0.0	63.1	0.2	24.3	15
	1 and 2	31.3	0.0	74.2	27.0	31.7	13
	1 and 3	29.1	0.0	72.7	17.7	31.2	14
	2 and 3	38.4	0.0	84.7	44.8	33.4	15
Eschar	1,2 and 3	24.5	0.0	85.8	0.0	37.7	9
	1 and 2	37.4	0.0	90.5	0.0	46.7	7
	1 and 3	26.5	0.0	90.8	0.0	40.8	9
	2 and 3	48.5	0.0	91.4	55.4	34.4	8

Table 8
Tissue agreement for Set 3 images

Tissue Types	Agreement Between Clinicians	Mean	Min	Max	Med	Std Dev	of Img
Granul	2 & 3	42.7	0.0	93.9	51.1	34.6	65
Slough	2 & 3	15.9	0.0	90.1	0.0	27.3	42
Eschar	2 & 3	25.0	0.0	92.3	0.0	34.6	30



Fig. 6. Two examples of very low agreement between 2 clinicians.

Table 9
Tissue percentage estimation.

Sets	Clinicians	Mean	Min	Max	Med	Std Dev	of Img
Set 2	1	23.3	7.6	51.7	19.4	12.8	14
	2	22.8	1.2	73.2	19.3	20.4	18
	3	19.5	0.7	48.4	16.4	14.8	16
Set 3	2	28.8	0.2	160.0	18.6	31.0	46
	3	25.4	0.1	132.7	16.3	27.4	53

that even the most experienced of clinicians are having trouble estimating the tissue percentages, which is an important piece of information required in wound documentation.

The results presented in this section show that wound segmentation and characterization are complicated processes, where even the most experienced clinicians have different opinions regarding wound boundaries and the type of tissues involved. The next section will discuss the results of the computer segmentation, and we will demonstrate that the proposed segmentation algorithm based on a probability map can be as good as the clinicians' consensus ground truth.

5.2. Segmentation and measurement accuracy

We carried out both qualitative and quantitative evaluations of the algorithm performance and these results will be presented in the next two subsections.

5.2.1. Qualitative evaluation

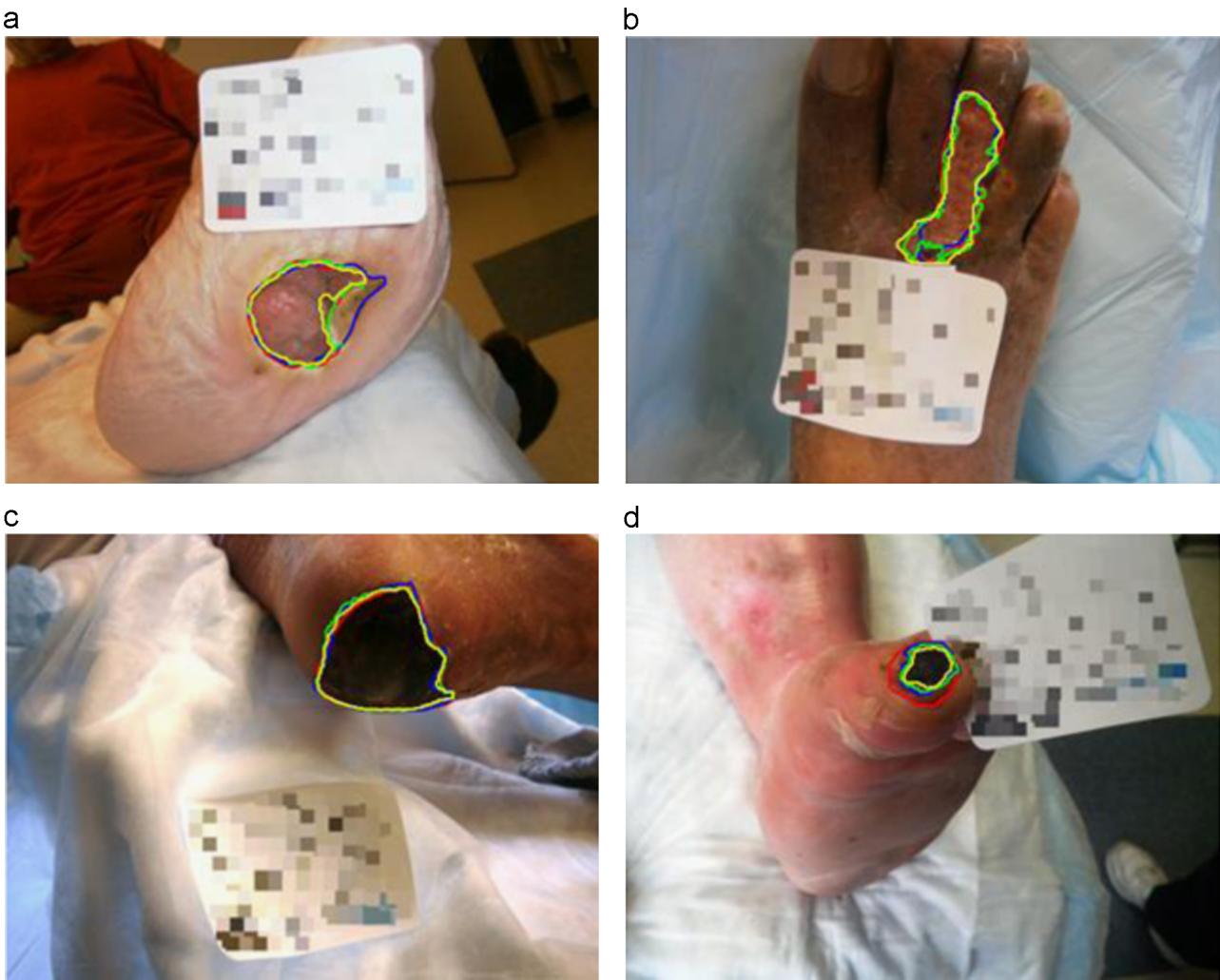
First, the performance of the segmentation algorithm was evaluated qualitatively. Fig. 7 shows four examples of the results obtained using both segmentation methods. For the first case (Fig. 7(a), granulation), the accuracy is 91.3% and 77.7% compared to the ground truth by Clinicians 2 and 3, respectively using the optimal thresholding, and 83.6% and 71.2% using the region growing segmentation. The discrepancies between the results against the different ground truths are caused by the rather big difference in the wound boundaries created by the two clinicians. For the second case (Fig. 7(b), granulation), the accuracies for both segmentation methods against both clinicians' ground truths are all more than 90%.

For the third example (Fig. 7(c), eschar), the accuracies are all more than 80% except for the optimal thresholding result against Clinician 3, which is around 75%. Finally for case 4 (Fig. 7(d), eschar), the accuracies for the optimal thresholding are recorded as 58.6% and 86.3%, while the region growing scores were 39.4% and 62.2%. As in case 1, the two clinicians differed in defining the wound boundary, where one of them included some parts of healed tissues as well, lowering the accuracy percentages for both methods. The optimal thresholding method agrees well with Clinician 3 with 86.3% accuracy, although the region growing approach seems to have missed some boundary pixels. The small size of the wound also contributes to further lower the accuracy of this particular wound image, due to the 'unforgiving' measurement metric used. Nevertheless, the four examples demonstrate that despite the complex nature of the wound boundary, the proposed algorithm is able to segment the wounds rather accurately.

5.2.2. Quantitative evaluation

Tables 10 and 11 present the overall segmentation accuracy using optimal thresholding and region growing approach respectively. Each table presents the results according to the different image sets as well as different clinicians' ground truths. It is observed that using optimal thresholding segmentation on the probability map provides slightly better overall results compared to using region growing. However, these differences diminish as the size of the dataset increases (i.e. Set 1 → Set 3), and the average accuracies become almost identical (74.0% vs. 74.2%). This trend is also true for individual clinician's agreements with the algorithm for different methods. Optimal thresholding is also more consistent than region growing as can be deduced by the lower standard deviation for all image sets. The overall average accuracy of 75% is very promising considering the level of agreement between the clinicians varies from 65% to 85%.

For performance comparison, we also run Sefexa image segmentation tool [27], which was developed based on the work by Kolesnik and Fexa [10–12], on the same sets of images, and the results are summarized in Table 12. Their method, like ours and unlike the other works discussed in Section 2, is not limited to images captured under controlled environment, not confined to the wound region, or designed for specific wound types only. Furthermore, besides supervised automatic mode, their method can also work on semi-automatic mode by requiring the user to provide samples of pixels belonging to wound and non-wound regions. These two factors make Sefexa, which is based on color and texture features, the most appropriate benchmark for our proposed method. Comparing the readings in Tables 10–12, both of our approaches outperform the Sefexa approach, which only records 68.8% average accuracy. Based on the standard deviation readings (10.5% for optimal thresholding, 13.1% for region growing and 17.0% for Sefexa overall), we can also deduce that our approach is more consistent. This is expected as Kolesnik and Fexa's approach depends heavily on the pixel samples to start the segmentation. While our approach requires the user to provide only an initial seed



Q2 Fig. 7. Selected segmentation results. For each case, the red and blue markings show the wound boundaries drawn by Clinicians 2 and 3, respectively, while the green and yellow markings show the segmentation obtained by the optimal thresholding and region growing approaches respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 10
Average segmentation (%) for optimal thresholding.

Sets	Consensus	Clinician 1	Clinician 2	Clinician 3	Average	Std. Dev.
Set 1	78.6	NA	NA	NA	78.6	8.0
Set 2	NA	79.6	77.4	73.5	76.8	9.8
Set 3	NA	NA	74.8	73.2	74.0	10.8
Overall	78.6	79.6	75.4	73.3	75.1	10.5

Table 11
Average segmentation (%) for region growing.

Sets	Consensus	Clinician 1	Clinician 2	Clinician 3	Average	Std. Dev.
Set 1	70.8	NA	NA	NA	70.8	14.3
Set 2	NA	77.1	75.7	73.6	75.4	10.8
Set 3	NA	NA	74.3	74.0	74.2	12.0
Overall	70.8	77.1	74.6	73.9	74.0	13.1

(i.e. a single click on an image), which is more convenient for the clinicians, the other method requires two sets of samples.

Table 13 shows the segmentation accuracy according to the different tissue types. Both approaches work best in segmenting granulation and eschar tissues, with lower accuracy for slough

Table 12
Average segmentation (%) for sefexa segmentation tool.

Sets	Consensus	Clinician 1	Clinician 2	Clinician 3	Average	Std. Dev.
Set 1	65.1	NA	NA	NA	65.1	22.8
Set 2	NA	78.9	78.3	80.4	79.2	10.4
Set 3	NA	NA	66.7	66.7	66.7	17.3
Overall	65.1	78.9	69.2	69.6	68.8	17.0

Table 13
Average segmentation (%) according to tissue types.

Tissues	Optimal Threshold	Region Growing
Granulation	76.2	75.3
Slough	63.3	63.9
Eschar	75.1	71.5

tissue. This is not surprising given the better delineated boundaries of granulation and eschar tissues. Slough tissues appear more sporadic, and also may be easily confused with other tissue types. This finding also agrees with the one reported by Wannous et al. [3]. **Table 14** compares the segmentation accuracies of the region growing approach between the proposed modified HSV color

Table 14

Performance comparison (%) between modified and original HSV for region growing.

Tissues	Modified HSV	Original HSV
Overall	74.0	62.9
Granulation	75.3	58.9
Slough	63.9	57.2
Eschar	71.5	66.8

space and the original HSV color space. Clearly, without modifying the HSV color space, the segmentation performance decreases considerably; highlighting the importance of our proposed modification. Without the modification, each of the overall wound segmentation as well as the granulation, slough and eschar tissues segmentation recorded a drop in accuracy between 5% and 15%. As expected, the granulation tissue segmentation benefits the most from our modified color space because better threshold is used to distinguish dark red (granulation) and black (eschar) tissues.

Optimal thresholding has much lower computational complexity compared to the region growing method. Region growing processes all the wound pixels, hence, the larger the image or the wound, the longer time is needed to complete processing all the pixels of interest. On average, to segment an image of size 768×1024 on 2.3 GHz Intel® Core™ i7 processor, optimal thresholding needed less than the second, while region growing required up to five seconds, depending on the wound size. Another issue to be considered when using the region growing approach for segmentation is the repeatability, i.e. the method should provide consistent segmentation results for different initial seeds. This is particularly even more challenging in our case as wound images tend to have "glossy" pixels within the granulation or slough area due to their wet nature. The optimal thresholding segmentation does not suffer from this problem, and thus is relatively more stable. Nevertheless, the proposed probability map approach, together with the mechanism to prevent premature stopping, is able to address this issue rather well.

6. Conclusion and future work

We have developed a method for the segmentation of wound images into granulation, slough and eschar regions and automatically carry out the measurements necessary for wound documentation. We proposed the red-yellow-black-white (RYKW) probability map as the platform for the region growing process in segmenting the three regions as well as the white label cards. Experiments were conducted on 80 wound images provided by The Ohio State University Wexner Medical Center. These images exhibited challenging characteristics with different types of wounds at different stages, typically pictured in a clinical setting with complicated backgrounds, some of which with similar characteristics to the color palette of the wounds or surrounding healthy skin. The analysis presented from the inter- and intra-reader variability experiment suggests that wound segmentation and characterization are a complicated process, where even the most experienced clinicians have different opinions regarding wound boundaries and the type of tissues involved.

Using the optimal thresholding approach, the proposed method achieves an overall accuracy of 75.1%, which is very promising considering that the average agreement between the clinicians is between 67.4 and 84.3%. The wound area, length and width measurements also give a promising accuracy of 75.0%, 87.0% and 85.0%, respectively. We have also demonstrated that the probability map approach, computed through a modified HSV color model, is a very promising method for use with many segmentation techniques to reliably segment wound images. Based on two simple segmentation methods, optimal thresholding and region growing, the overall

accuracy of around 75.1% has been observed. This suggests that the proposed RYKW map manages to identify the wound and its different tissues rather well, on par with the experts. Utilizing the RYKW map with a more advanced segmentation method can only further improve the accuracy of the segmentation, and is currently being worked on in our lab. The proposed method was also evaluated against other existing technique and experiment on the same sets of images shows much better performance for our proposed method.

We believe the proposed system will help wound experts immensely in the future. This early success could pave the way for a computer-assisted wound analysis software where the computer can segment the wounds reliably (with confirmation from the clinician) and provide a more accurate tissue characterization (as opposed to current clinicians' estimates), with possible extension into wound healing monitoring as well. With the tedious tasks of drawing the wound boundaries and populating the basic information on tissue characterization carried out by the computer, the clinicians will have more time in exercising their expertise in actual clinical work, thus achieving quality benchmarks for wound care as determined by the Center for Medicare Services.

It should be noted that the quality of the segmentation results as well as the resulting measurements depend on the quality of the input images. Unlike most of the previous work in this area, our work aimed at developing a solution that will work with actual, clinically captured images (all the images in this study were captured during routine clinical work and the personnel who captured them were not aware of software development). However, there is still the expectation that the images capture the wound in a reasonable manner; for example, if only a tiny portion of the wound is visible in the image, obviously, the segmentation will fail to properly capture the wound or its tissue components. Admittedly, human readers will run into the same challenge if asked to evaluate such images. Similarly, if the labels are not placed reasonably well, the absolute measurements may be skewed. Although our software can recognize some of the variations in the placements of cards, it cannot recover from severely distorted placement of cards. A ruler and color scale in the label cards can be easily included and these can be used to calibrate both size measurements and color variations, hence improving the overall accuracy. Other image acquisition issues include poor lighting and noise. While some of the images in our dataset do suffer from non-uniform lighting, noise and/or other artifacts (e.g. blurring in the images due to shaking the camera while taking the picture) to a certain degree, the proposed method performs rather well in handling these types of images. A future study needs to analyze the effect of such variations on the overall performance in a controlled manner.

The proposed algorithm has some limitations in segmenting and characterizing wounds on dark skins, especially when trying to identify eschar tissues or dark granulation tissues. In some rare instances, the color of Caucasian skins tend to be very red in appearance (in which the probability of red will be very high), hence segmenting fresh granulation tissues may not work on these images. We are exploring the possibility of incorporating edge and depth analysis into the current algorithm in order to address these problems, which could also potentially measure undermining wounds. In addition, work is currently under way to include images from other medical centers as well, and to further improve the segmentation accuracy by applying other segmentation techniques on the probability map. Automatic detection of the wounds, which would eliminate the need for the seed pixel by the user, is also under consideration. The proposed RYKW map is conveniently suited to achieve this by first identifying potential wound region throughout the entire image based on color information, before carrying out advanced analysis to filter the false positive regions. Finally, the ultimate goal of the wound software is not only to be able to characterize the wound at a single time, but also at multiple

1 time periods. By comparing the wound characteristics from the first
 2 visit to the second and subsequent visits, as well as taking into
 3 account the demographic information of the patient (age group,
 4 gender, ethnicity) and the type of ulcers (diabetic, venous, pres-
 5 sure), the healing rate can be estimated. This would be a significant
 6 breakthrough in wound healing management.

[10]

39

[11]

40

[12]

41

[13]

42

[14]

43

[15]

44

[16]

45

[17]

46

[18]

47

[19]

48

[20]

49

[21]

50

[22]

51

[23]

52

[24]

53

[25]

54

[26]

55

[27]

56

[28]

57

[29]

58

[30]

59

[31]

60

[32]

61

[33]

62

[34]

63

[35]

64

[36]

65

[37]

66

[38]

67

[39]

68

[40]

69

[41]

70

[42]

71

[43]

72

[44]

73

[45]

74

[46]

75

Conflict of interest

None.

[10]

48

[11]

49

[12]

50

[13]

51

[14]

52

[15]

53

[16]

54

[17]

55

[18]

56

[19]

57

[20]

58

[21]

59

[22]

60

[23]

61

[24]

62

[25]

63

[26]

64

[27]

65

[28]

66

[29]

67

[30]

68

[31]

69

[32]

70

[33]

71

[34]

72

[35]

73

[36]

74

[37]

75

References

[1]

[10]

53

[2]

[11]

54

[3]

[12]

55

[4]

[13]

56

[5]

[14]

57

[6]

[15]

58

[7]

[16]

59

[8]

[17]

60

[9]

[18]

61

[10]

[19]

62

[11]

[20]

63

[12]

[21]

64

[13]

[22]

65

[14]

[23]

66

[15]

[24]

67

[16]

[25]

68

[17]

[26]

69

[18]

[27]

70

[19]

[28]

71

[20]

[29]

72

[21]

[30]

73

[22]

[31]

74

[23]

[32]

75